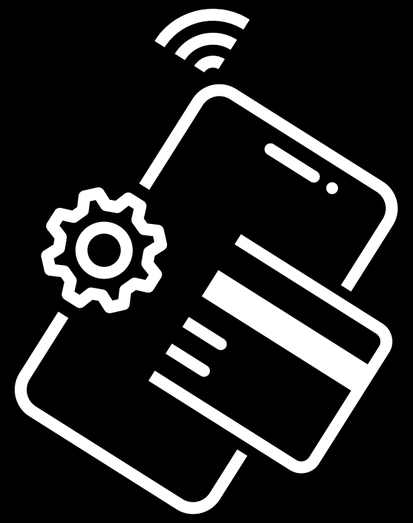
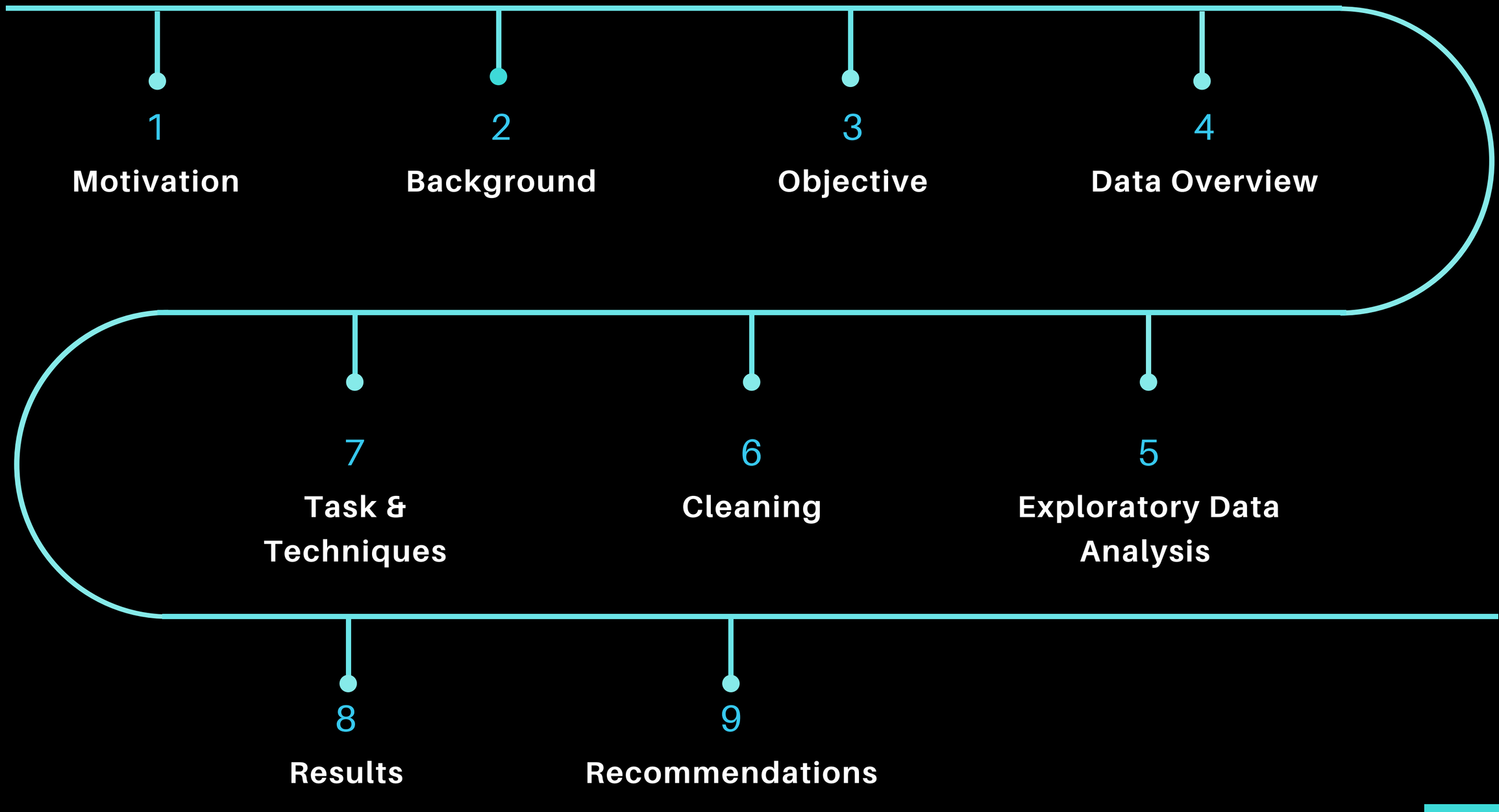
WHAT MAKES A CUSTOMER BUY A COSMETIC PRODUCT?



Amanda Sanelisiwe Nkomo
Mario Alcaraz

AGENDA


1
Motivation
2
Background
3
Objective
4
Data Overview
5
Exploratory Data Analysis
6
Cleaning
7
Task & Techniques
8
Results
9
Recommendations

What is the motivation behind this project?

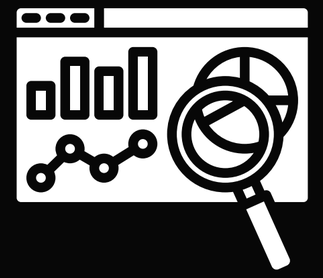
"Focusing on the customer makes a company more resilient."

Jeff Bezos CEO @Amazon

Every day thousands of people navigate online to purchase goods.
Gaining insights into the factors that drive consumer purchasing behavior is critical for companies to effectively market their products.

What is Marketing?

Sell/Distribute



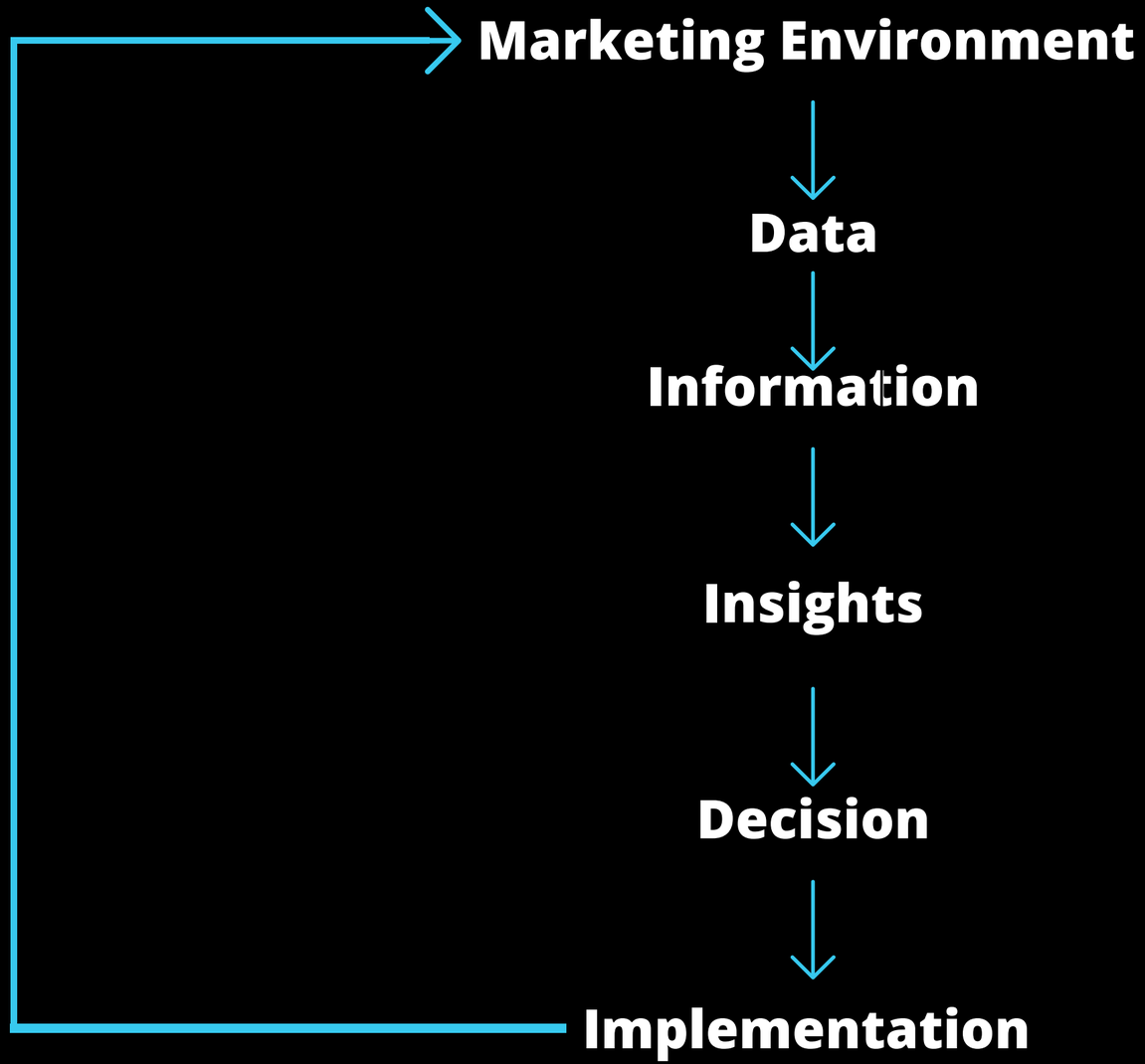
Market Research

MARKETING

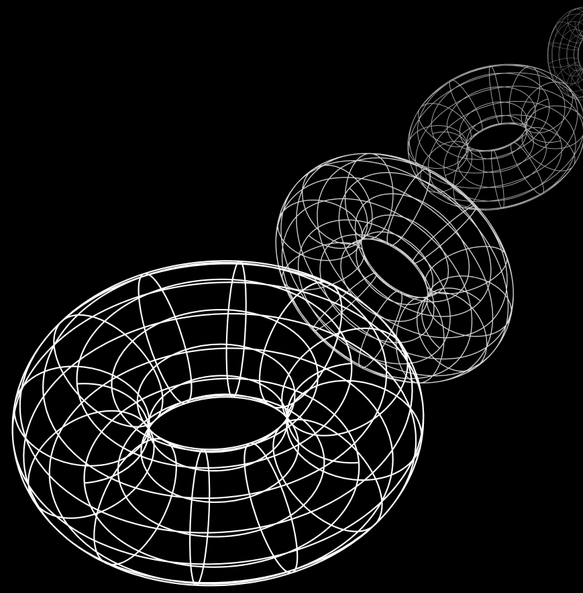
Promote

Advertise

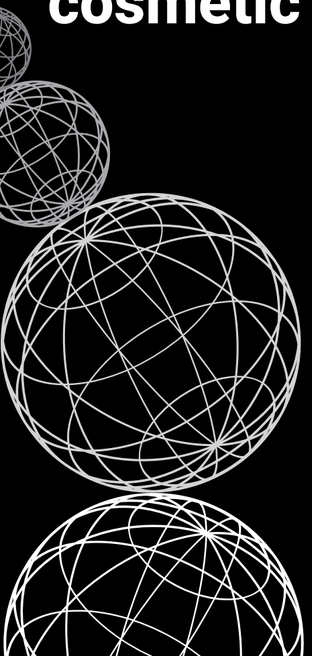
Marketing Engineering



OBJECTIVE



Apply the appropriate data mining techniques to help us accurately predict the sales of a cosmetic product and extract meaningful insights on consumer behavior for an effective marketing strategy.



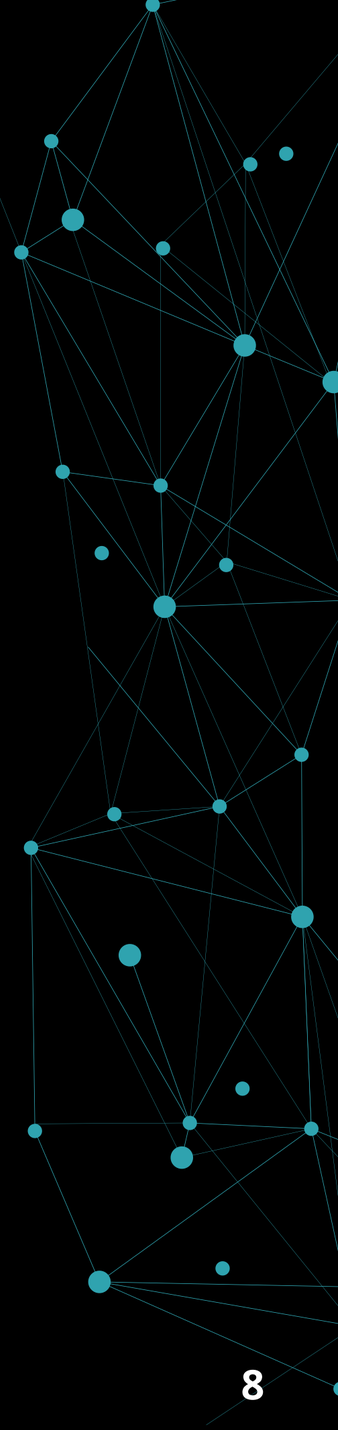
Our Data

E-commerce Cosmetic Store
Oct 2019.

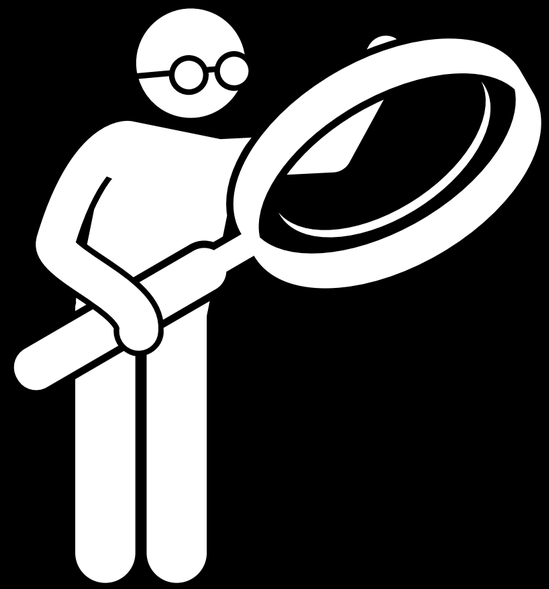
Observations: 4,102,283

Variables: 9

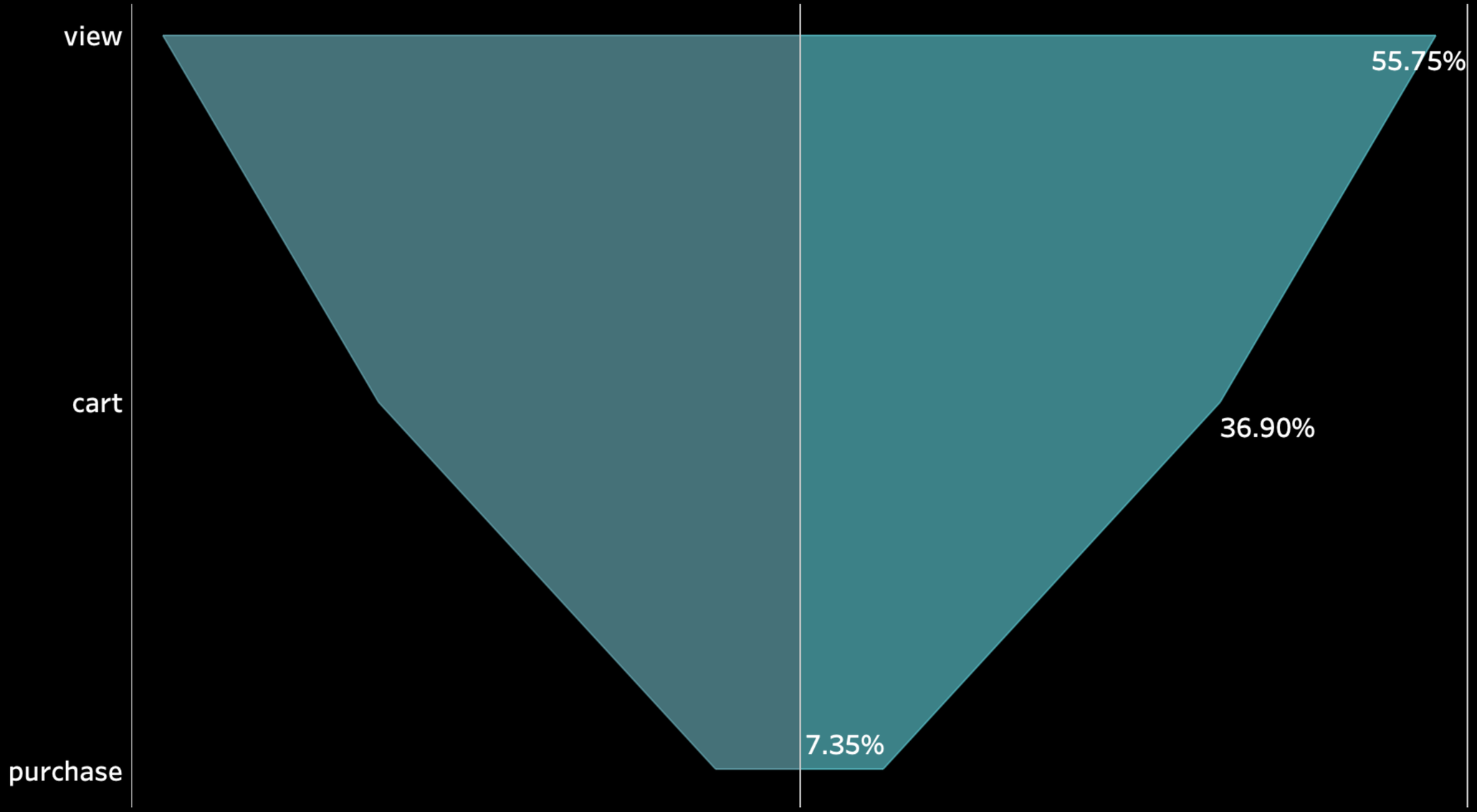
- Event Time
- Category Code
- User Session
- Event Type
- Brand
- Product ID
- Price
- Category ID
- User ID



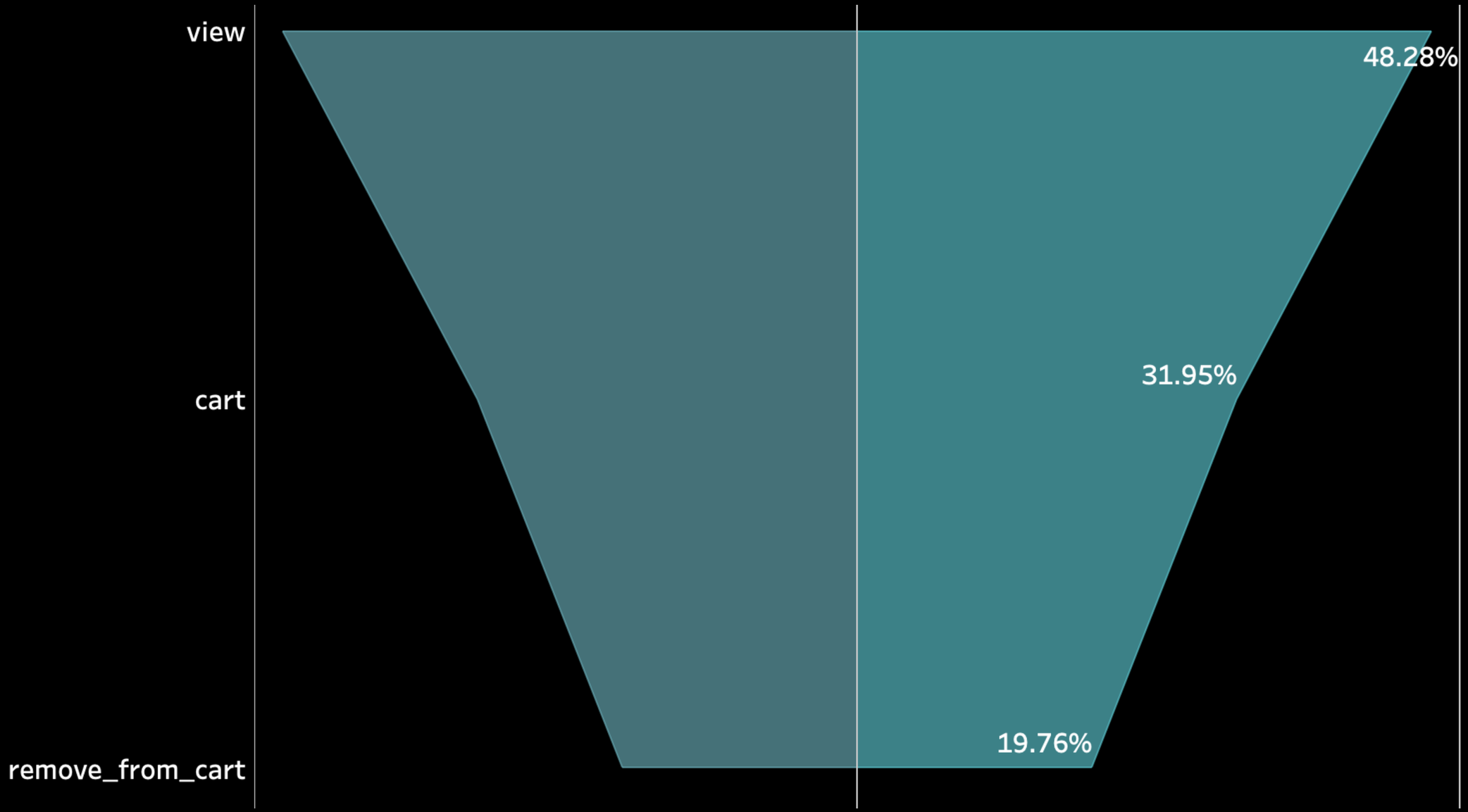
UNDERSTANDING THE DATA



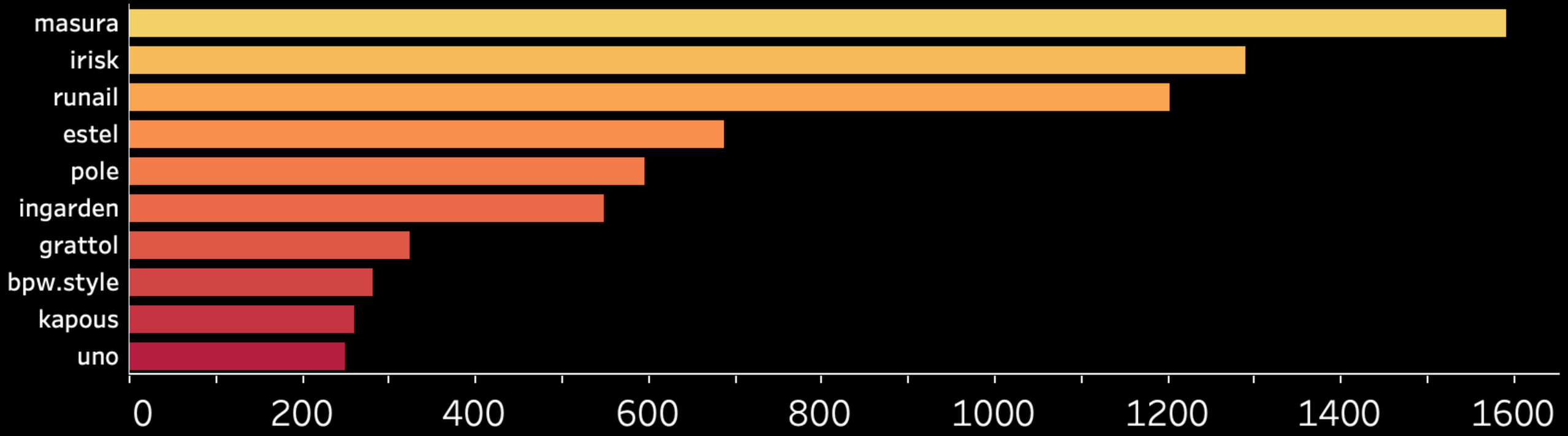
How many end up buying the product?



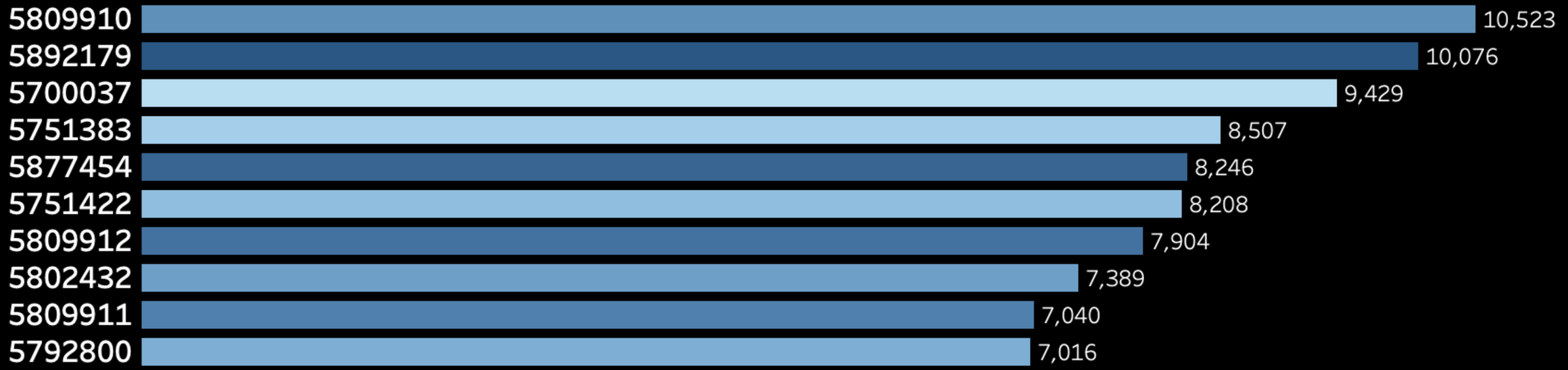
How many end up removing the product?



Which are the brands with most products?



Any Popular products?



Data Cleaning

Dimension Reduction

Category Code

Brand

User ID

Price

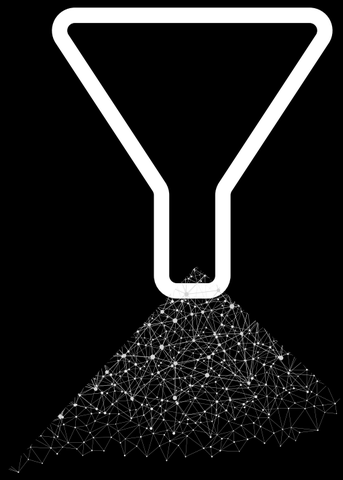
Event Time

User Session

Event Type

Product ID

Category ID



Price

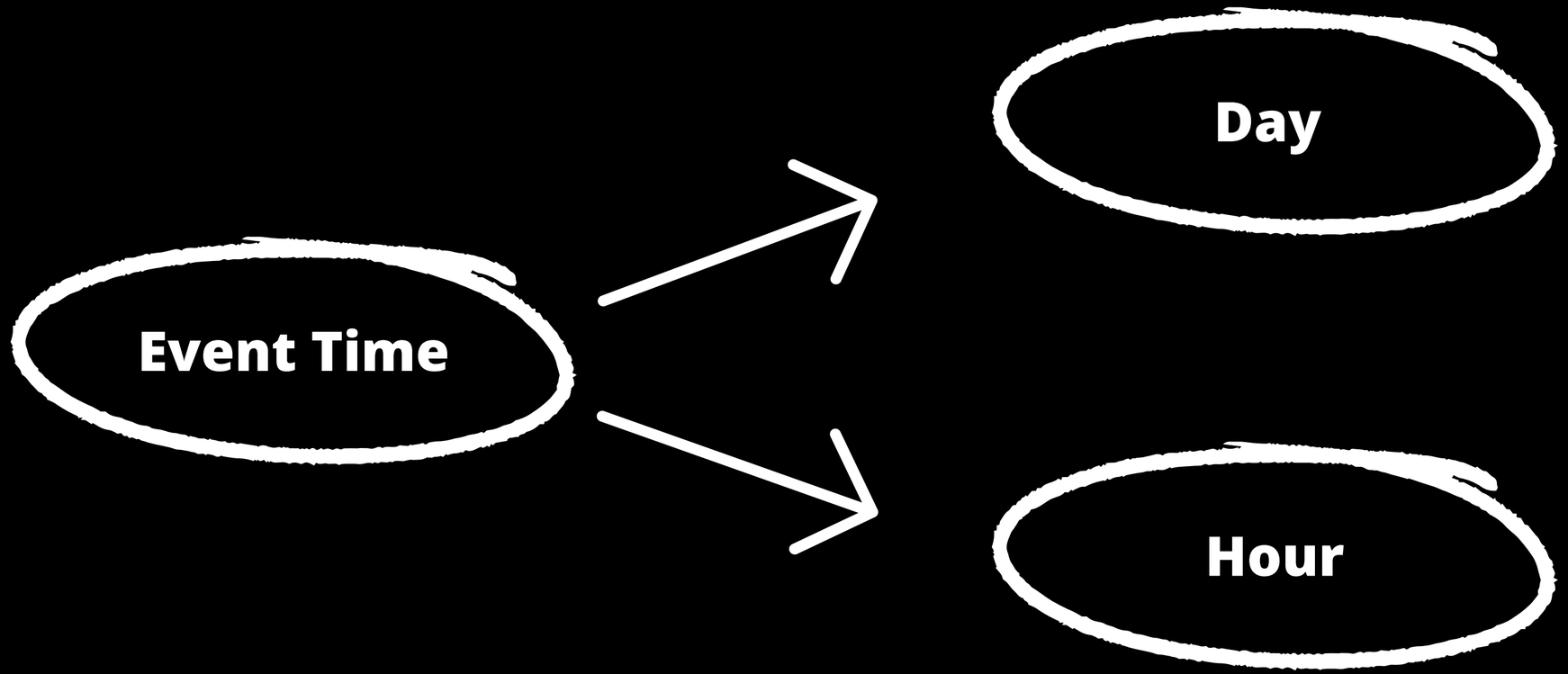
Event Type

Event Time

Product ID

Brand

Data Transformation



Generation of New Variables

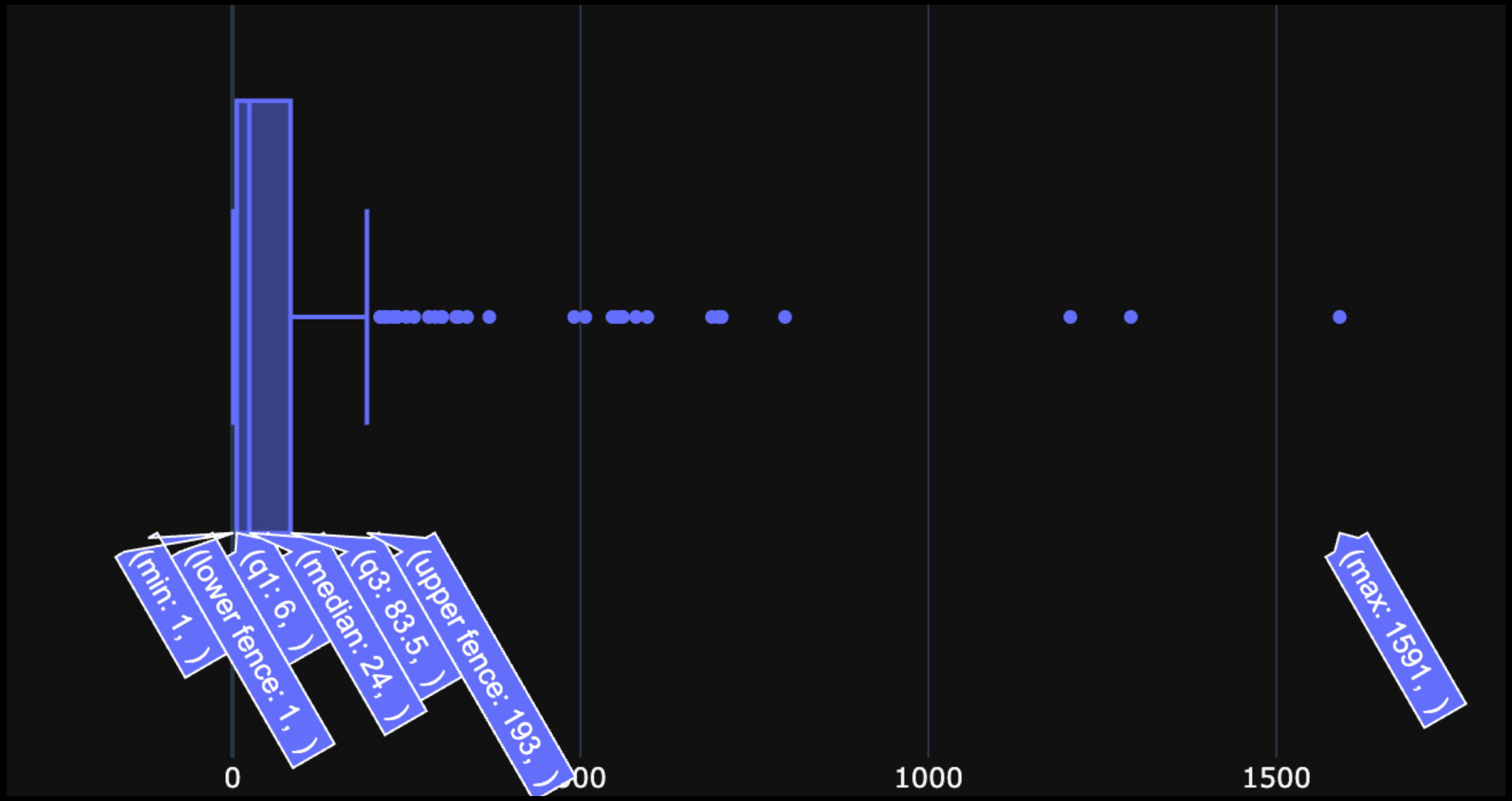
Big Brand

- **Count unique products per brand.**
- **Determine what makes it a big brand.**
- **Apply conditional function to choose 1 or 0 (1 = Big Brand, 0 = Not BIg Brand) .**

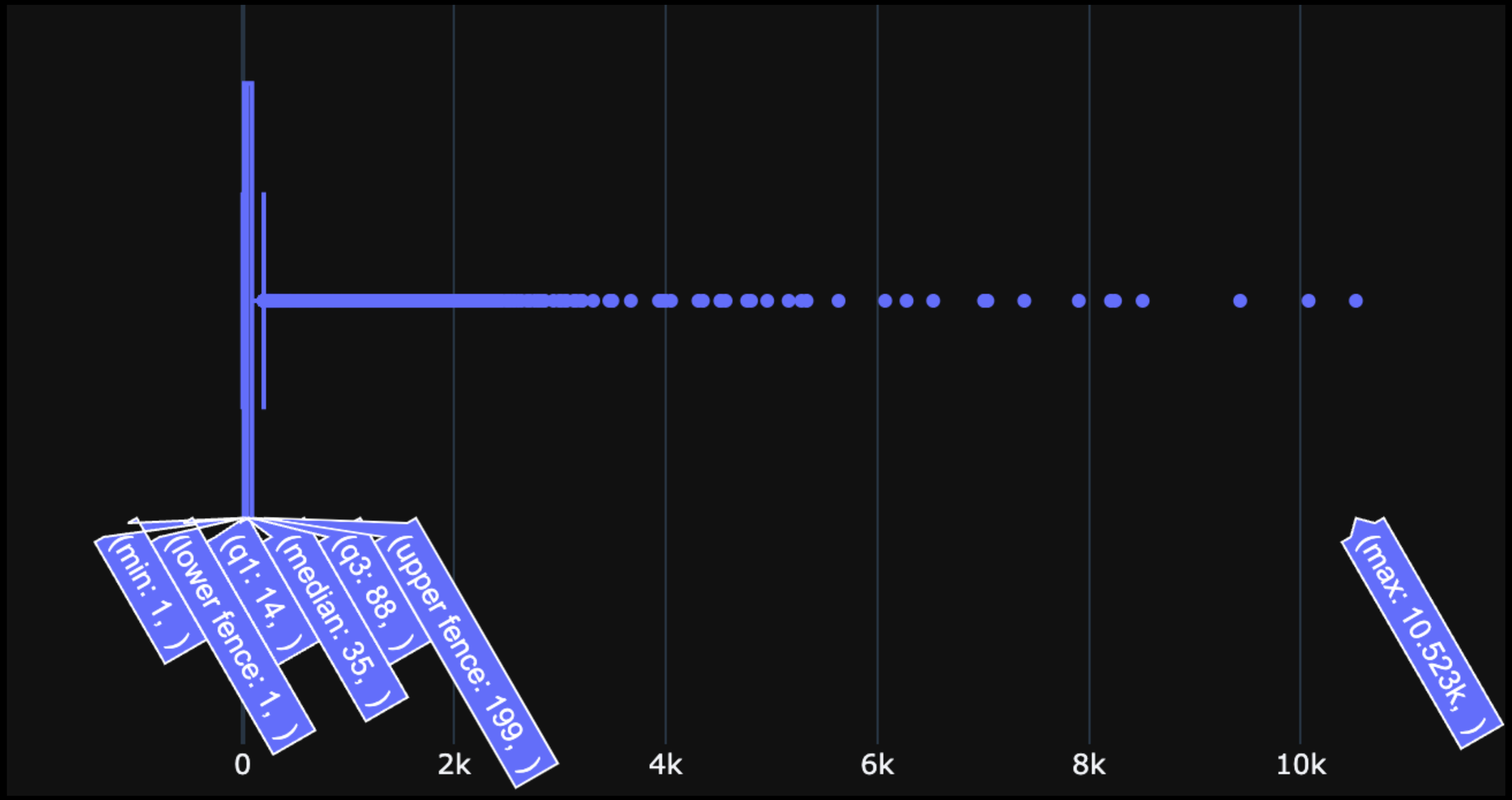
Popular Product

- **Count number of appereances for each product.**
- **Determine what makes it a popular product**
- **Apply conditional function to choose 1 or 0 (1 = Popular Product, 0 = Not Popular).**

Brand Size Distribution Analysis - Box Plot



Product Distribution Analysis - Box Plot



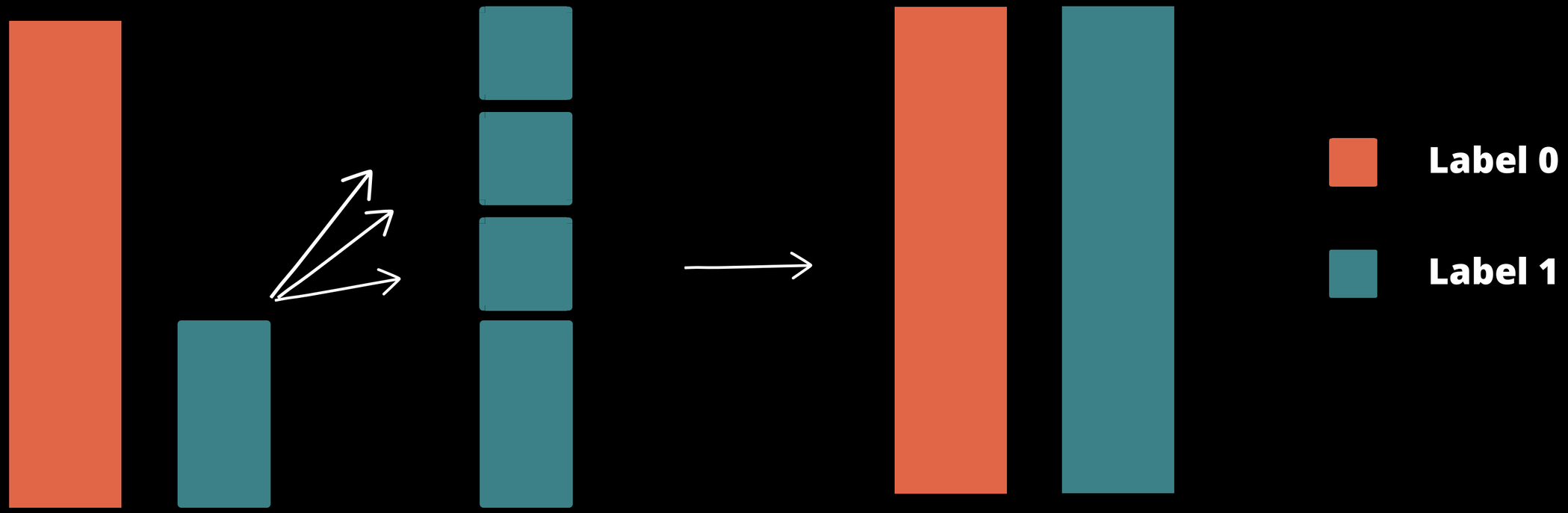
Clean Data

- **Price - Double**
- **Day - Int**
- **Hour - Int**
- **Brand Count - Int**
- **Product_count - Int**

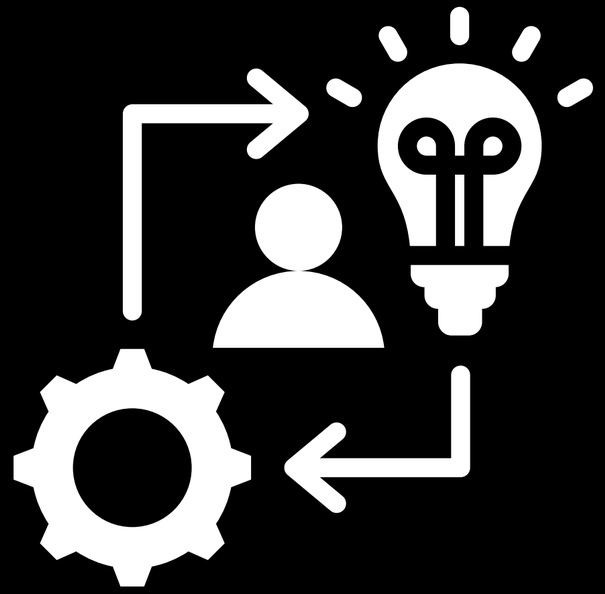
- **Big_brand - (1 or 0)**
- **Popular_Product - (1 or 0)**
- **Cart - (1 or 0)**
- **Purchase - (1 or 0)**
- **Remove from cart - (1 or 0)**

Data Balancing
Random Over-Sampling Examples (ROSE)

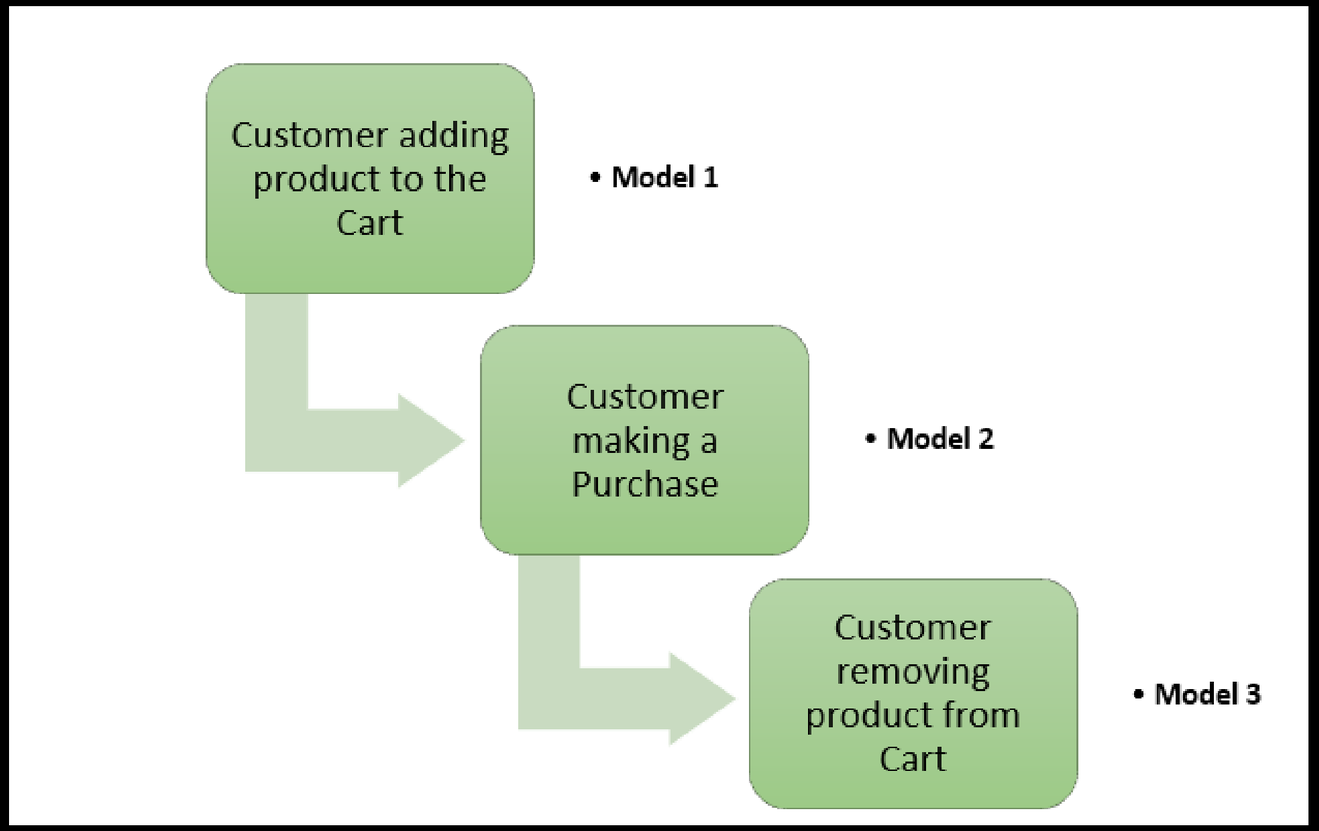
Bootstrap-based technique which aids the task of binary classification in the presence of rare classes.



Task & Techniques

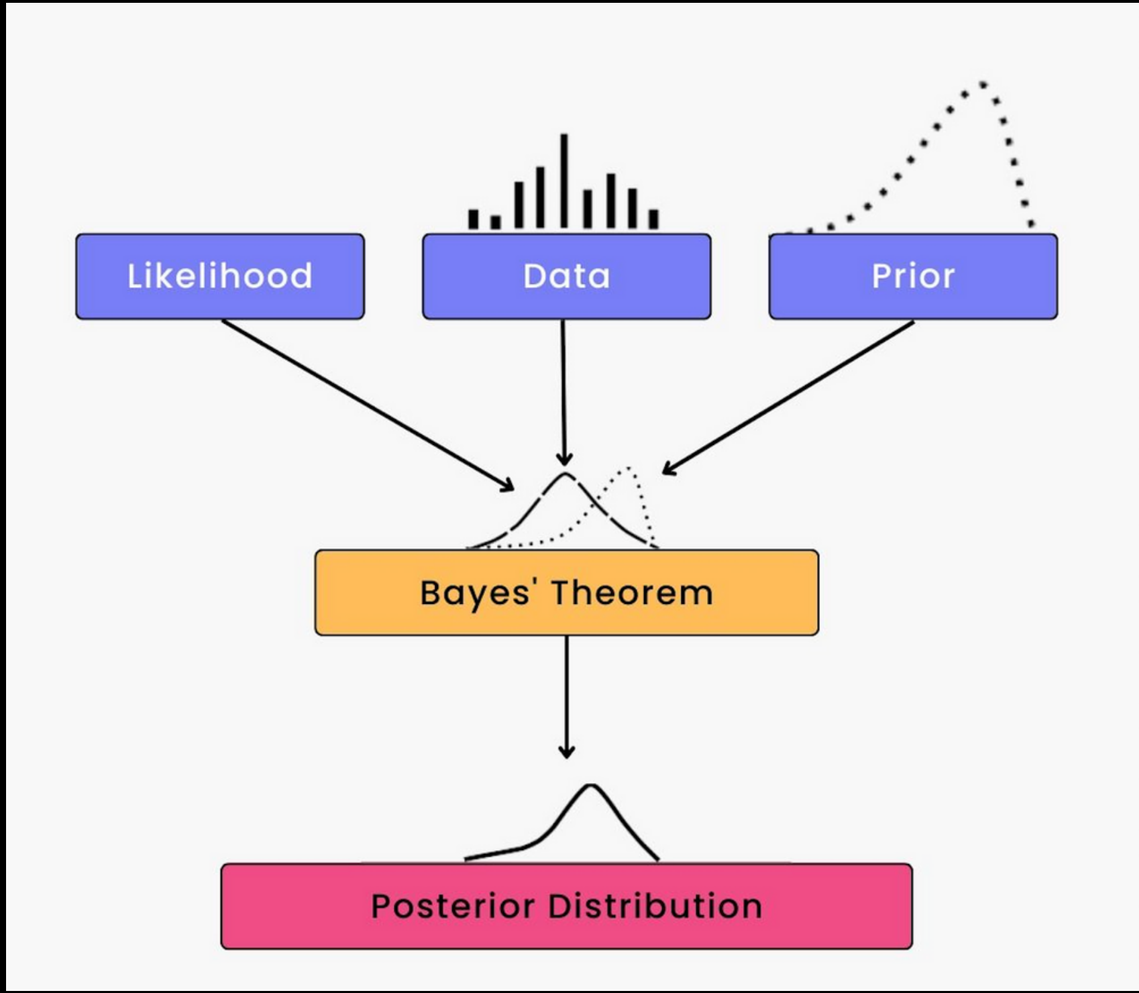


Logistic Regression



- Logistic regression models were used to analyse the relationship between binary dependent variables (event type) and the predictor variables (Big brand, popular brand, price, day, hour)
- The model coefficients and odds ratios quantify the strength of the relationship between the independent variables and the dependent variable.
- A confusion matrix can be used to evaluate the performance of the logistic regression model.
- Overall, logistic regression is a powerful tool for analysing customer behavior and predicting the likelihood of a customer making a purchase.

Naïve Bayes



- **Naive Bayes assumes that each feature is independent of the others, which means that we can calculate the probability of each feature independently.**
- **The model calculates the probability of each class for new data, using the probabilities of the attributes.**
- **Class with higher probability resulted is the one assigned**
- **Validate performance with new data and get confusion matrix.**

[1] Turing. (2022, March 11). Naive Bayes algorithm in ML: Simplifying classification problems

Added to Cart Predictive Model Results

ACTUAL

PREDICTED

	Not Added to Cart	Added to Cart
Not Added to Cart	44,159	15,376
Added to Cart	25,953	14,512

Accuracy: 58.67%

Sensitivity: 62.98%

Specificity: 48.55%

Purchased Predictive Model Results

ACTUAL

PREDICTED

	Not Purchased	Purchased
Not Purchased	84,225	4,917
Purchased	9,930	928

Accuracy: 85.15%

Sensitivity: 89.45%%

Specificity: 15.88%

Removed From Cart Predictive Model Results

ACTUAL

PREDICTED

	Not removed	Removed
Not Removed	62,921	12,520
Removed	18,785	5,774

Accuracy: 68.69%

Sensitivity: 77.01%

Specificity: 31.56%

Recommendations

- **Curb loss of sales from abandoned carts:** send cart abandonment notifications, discount coupons, simplifying the checkout process, and using retargeting advertisements to encourage customers to complete their purchase.

- **Make it easier to purchase:** Review adding to cart process, a quick button 'buy now' can improve customer experience and trim the journey from view, add, purchase to view purchase. A great example is Amazon.

- **Send beginning of the month and evening pre-order/save:** The later in the day and in the month, the less likely the item to be added to cart. This could be because people have spent their utilities later in the month and/or once they are at work they forget about adding to cart.

- **Promote popular products at end of month:** purchase predictive model showed that popularity of product and day are contribute positively to the item being bought.

- **Customer Lifetime Value (CLV):** Measure the potential long-term revenue that a customer can generate for an online business, we suggest using . Use models can to identify the most valuable customers and focus marketing efforts on retaining them.

Questions?

References

[1] Turing. (2022, March 11). Naive Bayes algorithm in ML: Simplifying classification problems. Naive Bayes Algorithm in ML: Simplifying Classification Problems. Retrieved May 2, 2023, from https://www.turing.com/kb/an-introduction-to-naive-bayes-algorithm-for-beginners

Appendix

R Notebook

Python Notebook